Document Imaging Strategy and Procedures for SFA Electronic Records Management System

June 15, 2001

TABLE OF CONTENTS

1. Introduction

1.1 Purpose

The purpose of this document is to provide the strategy, activities, roles, responsibilities, and procedures for the back-file conversion at the Department of Education's Office of Student Financial Assistance Programs (SFA). These tasks are part of the Electronic Records Management System (ERMS) project for SFA in which a large inventory of paper documents must be converted to digital format.

1.2 Document Organization

Section 1—**Introduction** describes the purpose of the document, includes a brief overview of the organization of the document, and lists key sources of information and acronyms used in the document.

Section 2**—Back-file Imaging Strategy** provides the strategy, activities, roles, and responsibilities for performing conversion of the back-file documents.

Section 3—**Back-File Imaging Procedures** provides the procedures for converting the back-file paper documents to electronic format and storing them into the ERMS.

Section 4—**Quality Control and Assurance** provides information on ensuring quality throughout the conversion process.

1.3 Information Sources

The following information sources were used in the development of this document:

- EDS. (2001, April 20). *Electronic Records Management Proposal Task Order #25 (Modification 01)*. [Proposal]
- Burke Consortium. (2001, February 6). *NAVSEA Corporate Document Management System (CDMS) Backfile Conversion Processing Guidelines For Conversion Imaging Partners*. [Document]
- The Rheinner Group. (2000). *Document Imaging: An Implementation Workbook*. [Document]. Hingham, MA: Author.
- The Rheinner Group. (2000). *Production Imaging*. [Document]. Hingham, MA: Author.

1.4 Acronyms

The following acronyms are used in this document:

DRCC	Document Receipt and Control Center
EDMS	electronic document management system
OPE ID	Office of Post-Secondary Education Identification
PDF	portable document format
POC	point of contact
SFA	Student Financial Assistance
SME	subject matter expert

2. Back-File Imaging Strategy

2.1 Overview

Back-file conversion consists of converting paper documents into electronic images with the associated indexing information. The document imaging process will be conducted and completed in a structured and controlled manner. The imaging process emphasizes security and control of the files during transport, storage, and processing to provide protection against destruction, unauthorized access, theft, and other security issues. In addition to security measures, the imaging process provides planning and management, conversion setup, document preparation, shipping, inventory tracking, scanning, indexing, release of the electronic documents onto unalterable media, and quality assurance and control.

Indexing information will be captured for each document and stored onto the selected media along with the document. The paper documents will be converted into searchable electronic text and images that can be efficiently stored and retrieved. The output format that will be used is Portable Document Format (PDF) Original Image With Hidden Text.

2.2 Back-File Strategy

The back-file conversion process will be conducted both on-site at SFA and off-site at the imaging partner's facilities. Some of the activities that will take place before the paper documents are converted to electronic files will be processed on-site. The remaining activities will be processed off-site. Detailed descriptions of the conversion activities are provided in the following sub-section. The files have been characterized into the following two types:

- Dormant—files that do not have re-certifications scheduled and are less likely to be requested by the customer during the imaging process
- Active— files that are scheduled for re-certification and likely to be requested by the customer during the imaging process.

These file types have been defined so that the conversion can be conducted in a manner that is optimal for continuing the normal workflow.

The back-file conversion will be conducted in three phases. Phase 1 will be the Pilot and will consist of a small subset of files from the Document Receipt and Control Center (DRCC) that are dormant and a small percentage of files that are active. The active files will be converted at the end of the Pilot phase when the ERMS is available to store and retrieve the documents. Phase 2 will involve conversion of the remaining dormant DRCC files. The files for Phase 2 will be processed immediately following Phase 1. Some overlap of Phase 1 and Phase 2 processing may occur to ensure that the flow of documents is maintained at the volume needed to complete the conversion within the targeted time frames. Phase 3 will consist of processing the remaining DRCC files that are active. This strategy of converting dormant files first will be employed to minimize the time that the DRCC will be without a given file.

The resources required to complete the back-file conversion project are categorized as Hardware, Software, Supplies and Personnel. The hardware and software required to inventory, track, scan, index/validate, and release/store the files will be determined and provided by the imaging partner. The boxes, labels, document separator pages, and other supplies needed for packing, shipping, and preparing the documents for conversion will be determined and provided by the imaging partner. Based on the expected conversion timeframe and volume, Mod Partner recommends the following for the personnel resources required to complete the conversion process in a timely manner:

- 1 SFA Subject Matter Expert (SME) / Point of Contact (POC),
- 1 Coordinator,
- 5 Document Analysts, and
- 2 On-Site Imaging Specialists.

Additional off-site imaging specialists will be provided at the imaging facilities to provide shipping and other off-site imaging services.

2.2.1 Conversion Activities

The activities involved in the successful conversion of the back-file paper documents are listed below.

- Planning and Management —Planning and management are key activities to ensure the success of the project. Detailed schedules, procedures, and logistics will be defined and documented for processing the material at SFA. Contingency plans will be defined, additional equipment will be secured as required, and production schedules will be adjusted to meet the defined time periods. Meetings with the customer, imaging partner and Mod Partner will take place to identify all details and necessary information. Planning for and management of quality will also be conducted.

- Conversion Setup —Setup involves the installation of the necessary hardware and software, setup of the tracking and indexing databases, and implementation of processes for conversion to begin.

- Pre-Preparation —Sort files to determine which files will be converted. Organize file folders into the appropriate order so that each file is ordered in the same sequence. Identify the metadata fields and field location for each document type and version within a document type.

- Catalog and Prepare Files for Shipping—Catalog and label files and boxes for tracking purposes as they move through the imaging process. Place files in boxes and prepare for shipping them to the imaging partner's site.

- Initial Shipping—Ship boxes to imaging partner locations for scanning, indexing, and other imaging tasks.

- Document Preparation—Prepare documents for scanning. This includes removing impediments, taping undersized documents to a standard size sheet of paper, and flagging documents.

- Indexing —Enter metadata into the index fields for search, retrieval, retention, and disposition functions within the ERM.

- Scanning—Convert hard copy documents into electronic images and produce PDF Original Image With Hidden Text format.

- Indexing Validation—Validate the accuracy of the metadata entered into the index fields.

- Release/Store—Release the converted images and associated indexes into the designated repository or temporary storage devices.

- Re-assemble Files—Re-assemble documents/files according to specifications so that they can be returned to SFA and/or disposed or stored properly.

- Return Shipping—Ship the boxes for return to the appropriate SFA organization or to the appropriate storage location.

- Quality Assurance and Control —Perform quality control checks on document images and metadata to ensure accurate conversion to electronic format.

2.2.2 Roles and Responsibilities

The team of personnel resources required to complete the conversion consists of SFA personnel, the Mod Partner resources, and the selected imaging partner who provides the electronic imaging specialists. The roles and responsibilities of the team are listed below.

SFA SME/POC:

- Provide document/file and processing information and requirements

- Provide answers to questions, assistance with problem resolution, and pre-prepares documents

- Provide personnel to release files for shipping, accept return of files, and sign off on manifest/tracking documents

- Provide timely access to the documents during regular business hours

Coordinator:

- Provide coordination, oversight, and management throughout the conversion process
- Monitor and validate activities
- Sign off on manifest/tracking documents
- Assist in problem resolution
- Act as the liaison between customer and the imaging specialists
- Verify imaging results and ensure quality

Document Analysts:

- Pre-prepare documents for conversion
- Validate and test samples of the converted output for accuracy and quality

On-site Imaging Specialists:

- Assist the customer in the pre- preparation activities
- Place bar code page separators during pre-preparation
- Scan barcodes into tracking and inventory system

Off-site Imaging Specialists:

- Ship documents to and from the customer site
- Perform conversion activities to convert, index, validate, and release images and indexing information.

The SFA/DRCC office will provide the SFA SME/POC and document analysts for the conversion effort. Mod Partner will serve in the role of Coordinator. The imaging partner will provide the on-site and off-site imaging specialists. SFA will be responsible for providing space

on-site for the conversion team to perform their tasks. Space is required to setup the tracking and inventory database and for packing/boxing, loading, and unloading.

2.2.3 Phases

2.2.3.1 Phase 1 – Pilot

Timeframe: 2.5 months

The Pilot phase will consist of converting a small sample of documents for a specified group of approximately 150 schools. This sample of documents equates to approximately 150,000 to 300,000 pages. Additional schools that are currently active and scheduled for reviews, audits, analysis, etc. will be identified for processing during the Pilot phase. These schools will be added to the process once the ERMS repository is available to store the documents. The purpose for adding these schools is to test the imaging process and the ERMS using active files.

2.2.3.2 Phase 2 – Dormant DRCC Files

Timeframe: 3 months

The files for this phase will be processed after the completion of phase 1. These files will not be accessed via the repository until the available for access. Therefore, a means for providing the files via other electronic media will be available.

2.2.3.3 Phase 3 – Active DRCC Files

Timeframe: 3 months

The remaining files that are active and have a high probability of being required during the back-file conversion will be processed during this phase. This phase will start when phase 2 is complete and the repository is available for access. Once the files in this phase are picked up for shipment they will be converted and accessible via the repository within four to seven days. If a file is needed before it is available in the repository, a means for providing the file via other electronic media will be available. To keep pace with the target date for the overall completion, approximately 3 million pages will need to be processed within a three-month period.

3. Back-File Imaging Procedures

3.1 Overview

The following sections provide the descriptions of the procedures required for completing the back-file conversion. Throughout the conversion activities described, Mod Partner will coordinate and oversee the efforts, ensure accuracy and completion, schedule tasks, facilitate imaging partner/customer communication and problem resolution, ensure adherence to the schedule, communicate task due and time frames, and provide status reports. Throughout the procedure descriptions the work products are referred to in various terms, which are files, folders, documents, batches, and shipments. A file is defined as an entire school record, which is identified by Office of Post Secondary Education Identification (OPE ID). Folders are the units within the files that contain the documents for that school. The documents are the collections of writings contained within the folders and provide information on a school. A batch is a box of documents, which may contain a subset of a school's information or information for more than one school in one box. A shipment is a collection of boxes shipped together on the same day. Throughout the entire process the documents are tracked to ensure data integrity, document security, and to assist in assuring and controlling the quality of the process.

3.2 Pre-Preparation

Within the DRCC the file level consists of all the folders and documents within the folders for a specific school. Folders are the subsections within a file. The subsections are based on cutoff dates and the following categories: Eligibility and Certification, Financial Statements, Audits and Program Review. The document analysts will go through all of the files and folders to determine which folders within each file will be converted. Folders that will not be converted will be boxed and archived. Folders that will be converted will be left on the shelves for further processing.

The SFA SME will determine the prioritized order for each file and the exact order for the documents within each file. The pre-preparation team will document that order. The document analysts and imaging specialists will work together to place each file folder in the defined order. The coordinator will document the order as defined, ensure that the order is adhered to, ensure that problems and issues are identified, raised, documented, and resolved regarding the organization of the files and identification of the metadata.

While the files are being organized, the metadata fields and field locations on the document for each document type and for each variation of the different document types will be identified. The coordinator, SFA SME and the document analysts will work together in documenting the metadata fields and field locations for the documents. The team will work together in identifying problems and issues with the metadata such as missing values, illegible print, inconsistencies, and other issues. The coordinator will ensure that problems and issues are identified, raised, documented, and resolved regarding the organization of the files and identification of the metadata.

While the files are being organized, the imaging specialist will affix a barcode separator page at the beginning of each file folder and/or first page of each document. If the individual files already have barcodes, then the existing barcodes may be used. Each barcode will be associated with a unique number used to identify the file and metadata for that file. The barcodes will be scanned prior to shipping with hand-held barcode scanners and stored into a database. This will identify each file and document. This barcode will be matched with the corresponding metadata during conversion. The numbering scheme for the file level (school file) will be the OPE ID number, which is a unique eight-digit number previously assigned to each school by SFA.

The imaging specialist will also identify the files and documents that have exceptions to the defined conversion standards and require additional costs to process. These exceptions along with the associated costs will have to be documented and approved via signature by the designated authorities prior to shipping.

3.3 Cataloging/Shipping Preparation

Once the pre-preparation is completed and just before it is time to ship the documents for conversion, the coordinator will notify the conversion team of the time to begin the Cataloging and Shipping Preparation. Prior to starting the activity SFA will provide a location and make all of the files available for this activity. The imaging specialists will identify each box with a barcode number and affix a barcode to each box. This box barcode number is part of is the batch tracking number. The batch tracking number will be used to identify what documents are inside the box and where the box is located while being loaded, in transport, during delivery to the conversion facility, and when they submitted for conversion. It serves as a check-in/checkout number for processing. Therefore if a file is inadvertently separated from its batch, it can be quickly identified and returned to its proper storage location or position in the conversion process.

The boxes will be uniform in size and dimension to conform to the storage facility requirements, and the document handling equipment requirements used to process it. Once each box/batch is labeled with a batch tracking number. The imaging specialists will scan each box's barcode into the database prior to loading it with files. Each file will be added to the batch number by scanning its folder and document barcode as they are placed in the box. The files will be packed into boxes according to the defined processing order and priority. A manifest will be printed that provides an inventory of the batches and files that are to be released for shipping.

The box barcode numbering scheme will consist of a unique, sequential number for each box. The group of boxes shipped on a given date will be referred to as a shipment. Each shipment will be given a unique, sequential number in the order that the shipment occurs. Each box represents a batch within a shipment. Each batch will be given an identification number using the following numbering scheme.

3.4 Shipping

3.4.1 Initial Shipping

For security, data integrity, and efficiency purposes Mod Partner recommends the shipping process provided below. Shipping will occur weekly and will be scheduled in advance. The team will be given the shipping schedule and the coordinator will endure that everyone is aware of the scheduled dates so that they may prepare accordingly. To ensure data integrity each shipment must remain together throughout the entire process. The imaging partner will define the predetermined route for shipping and provide the route information to Mod Partner and SFA/DRCC. The manifest must be reviewed and approved by signing the document to release the files for shipment. The imaging partner will then move the boxes to the loading area and onto the truck once each box has been loaded with the files.

As each box leaves the facility the affixed bar code will be scanned to indicate that it is in transport. Transport vehicles will travel along prescribed routes. All transport vehicles will be equipped with mobile phones and personnel with pagers. Transport personnel will be required to sign for the load at the point of origin and check-in upon arrival at the conversion facility. Loads will be inventoried upon arrival to ensure all material is present. Transport vehicles will be manned by a minimum of two personnel for all major transfers, and our team's management staff directly supervises them. If a transport vehicle does not arrive at the prescribed time and the transport personnel cannot be contacted, a second team is dispatched to locate the vehicle and the customer is notified. If the team fails to locate the missing vehicle and personnel along the prescribed route, the customer and police will be notified after an appropriate waiting period has expired. This way the possibility of lost or misdirected material will be minimized and potential exposure limited.

3.4.2 Return Shipping

Once the files are converted and reassembled, return shipping to return the files to SFA will be done in a manner similar to initial shipping. The returned boxes will contain the same files that were in the boxes during initial shipping. All documents within a shipment must be returned together with the files they accompanied during pickup. There will also be a newly generated manifest sheet for the return shipment and each box will contain a newly generated box content sheet. Upon return of the documents, manifest that is returned with the shipment must match the original manifest sent out with the shipment during pickup. The document counts for each box, the number of boxes, and box content sheets for the return shipment must contain the same information and values as that which was defined for the initial shipment. After verifying the above, the conversion team will review the return manifest and sign off on it as acceptance of the returned files.

When the documents are returned they will be sealed in boxes and placed in storage. The hard copy documents will no longer be used for circulation. The documents will be available in the ERM repository and/or through other electronic media if required.

3.5 Conversion

Document conversion consists of a series of steps that are designed to ensure quality. They consist of the following: document capture, enhancement, recognition, indexing, quality assurance and release to the ERMS. The conversion activities will take place off-site at the imaging facilities and consist of the following activities:

1. Document Preparation
2. Indexing
3. Scanning
4. Index Validation
5. Release/Store
6. Re-assemble Files

3.5.1 Document Preparation

During document preparation the documents will be made ready for scanning and prepared for initial indexing. Proper preparation will help ensure that the imaging operators can effectively process the material. The batch definition file determines how the document will be captured, processed, indexed, and exported. The batch definitions will be setup by the imaging partner's operations manager and are based on the customer's requirements, which will be determined prior to the beginning of each job. Since the operations manager defines the process and maintains configuration control over changes to the process, consistency is promoted.

A batch order will be prepared by the operations manager that indicates what material will be prepared for indexing and scanning, its location, instructions specific to the customer requirements for preparation work, and scanning instruction for the imaging operators. When the imaging operators receive the work order, the document batches will be retrieved from the warehouse and checked out to the processing area. The document preparation personnel will prepare the material in accordance with the batch instruction sheet. This may consist of but is not limited to following:

- Remove the material from the boxes
- Identify any hazardous conditions (lead staples, molds, infestation problems)
- Check the file numbers against the batch
- Ensure the order of operation
- Remove documents from the file folder

- Remove all additional binding materials (staples, paper clips, file clips, or other binding devices)
- Place used materials in the proper sort bins
- Evaluate the condition of the paper to ensure it is within acceptable tolerance levels for the scanners and identifying any special handling requirements
- Photocopy exception pages (torn, taped, or otherwise damaged)
- Remove edges or binding if required
- Jog the material
- Place the prepared material on the document cart
- Document the material progression
- Deliver the documents to the next operator for processing

3.5.2 Indexing

Prior to scanning the hard copy documents to electronic form, the index information for each file and document will be entered into the indexing system.

3.5.3 Scanning

Documents capture is the scanning of documents according to the instructions provided on the work order. The scanned files will be converted to PDF Original Image With Hidden Text. The work order identifies the proper machine setup for the scanning operator, what batch definition to use, and the filename or accession number for processing.

3.5.4 Index Validation

Once the documents are scanned, the index information is then matched with the barcodes and associated to the correct document. The indexing information is validated.

3.5.5 Release/Store

This activity must support the tracking system so that the documents can be tracked from the start of the process up to and including this step. As stated previously, DVD ROM will be the unalterable media used to store the output for the converted files. A minimum of 3 copies of the documents on DVD ROM is recommended. Naming and labeling conventions must be defined

and documented. These conventions will be used for naming the files and discs and for labeling the discs once they are created. The output must be consistent with these conventions.

3.5.6 Re-Assemble Files

Once the files are converted, indexed, and released to storage, the hard copy files will be re-assembled and placed back into the boxes as loose pages in the order they were received. Any bindings, staples, etc. that are removed will not be replaced. When re-assembling the boxes, all documents must be packed in the same order and with the same documents with which they arrived. The imaging specialists must make sure that for each box the contents and counts for the outgoing shipment match those in the incoming shipment. Each box must contain a newly generated content sheet for the return shipment. In addition, each return shipment must contain a newly generated manifest for the shipment.

4. Quality Assurance and Control

Quality Assurance and Control involves checks throughout the process to ensure readability of documents and accuracy of indexing information. Quality Assurance and Control is represented as the last step in the back-file conversion process; however, quality control mechanism must be incorporated in the entire conversion process to ensure that quality output is achieved. In essence quality assurance is a part of each step and not just a single step in the process. To verify the quality of the output acceptance criteria must be established and test procedures must be defined and executed. The following sections provide minimal guidelines for incorporating quality throughout the back-file conversion process. For each step in the process, applicable guidelines are provided.

Planning and Management

Quality must be incorporated in the planning and management for the project. The schedules must incorporate time for tasks to implement quality procedures and verify the quality of the output. Quality measures must be incorporated in the planning of procedures. Contingency, Disaster Recovery, Problem Solving, Quality Assurance and Control and Test plans must be developed and managed. Project leads and managers must ensure that quality is planned for and carried out in all steps and phases.

Conversion Setup

During the conversion setup, a tracking system, using barcodes, to track each document through the system should be established. This enables the imaging partner to ensure that each document is processed through each step. It also helps in keeping track of the counts for the documents that have been processed.

In addition to the tracking system, an indexing system must also be established so that the correct metadata can be entered and verified. These tools are used throughout the process to aid in quality assurance and control.

Pre-Preparation

During pre-preparation the indexing guide sheets for each document type and version must be developed for use by the imaging specialists during the entry of the metadata values. These sheets must be checked for accuracy before they are forwarded on to the next activity. This helps ensure that the correct metadata is entered for each document.

Barcode separator sheets must be generated and placed in front of each document so that every document can be tracked through the entire process. This also enables the imaging specialist to associate a file with the correct set of metadata.

Cataloging and Shipping Preparation

All documents must be entered into the tracking system during cataloging and shipping preparation. The tracking system will be used to determine the boxes' content, number of boxes and documents per shipment, and the number of documents per box. This information will be printed on box content and shipment manifest sheets for later use in verifying shipment content.

Initial Shipping

Upon arrival, each shipment must be inventoried and verified against the box content sheets included in each box and against the manifest sheet for the shipment. This quality check helps ensure that what was sent out arrived as expected at the imaging facility.

Document Preparation

During preparation the file numbers will be checked against the box content sheet to ensure the proper processing order. Batch definition sheets will be developed and used to ensure that the documents are processed correctly.

Indexing

The indexing guide sheets will be used to enter the metadata correctly for each document. Batch definitions will provide additional instructions as required for the indexing.

Scanning

During scanning the electronic document will be viewed for quality. If any scanned documents are found to be skewed, double fed or have overall poor quality, the imaging specialists will re-scan the document or page. Enhancement software will also be available to remove unwanted marks on a document and improve upon poor quality images.

Index Validation

During index validation the imaging specialists will check the metadata stored electronically in the index fields by comparing the values against the data provided on the hard copy. The specialists will ensure that the correct document is referenced in both and verify that the values are the same. If the values differ, then corrections will be made appropriately. While viewing the documents to verify the index values, the image quality will be checked again.

Release/Store

During the release/store activity content sheets will be generated for each DVD, specifying which documents are stored on it. The content sheets will be provided electronically on the corresponding DVD and in hard copy format. Naming and labeling conventions must be established for naming the files on the DVD and for providing electronic and hard copy labels for the DVD. When storing documents on the DVD and labeling it, the labeling and naming conventions must be consistent throughout each document and DVD to aid in the tracking and verification processes.

Re-assemble Files

Re-assembly of the files and boxes must be completed according to the original order and content. The original content sheets and manifests must be used as guides for re-packing the boxes and preparing the shipment for return. The document, file, and box barcodes should be scanned again during re-assembly. New box content and manifest sheets should also be generated to compare against the original sheets and ensure the outgoing shipment contents match the incoming shipment contents. In addition, the documents and boxes must be counted

while re-assembling to make sure that the count for the documents and boxes being shipped out matches the counts for the incoming shipment.

Return Shipping

Upon return of the shipment the SFA SME/POC and the coordinator should verify that the box content and shipment manifest sheets match between the initial and return shipment. Box counts should also be verified manually. Upon verifying this information the manifest sheet should be signed off in acceptance of the shipment. These quality control measures help ensure that the documents that were sent out completed the entire process and were returned to SFA.

Quality Assurance and Control

Quality Assurance and Control will culminate in the testing and verification of the output. A sampling of the each box will be taken and verified for quality and accuracy. It is recommended that the sample size be no less than 3% of the documents in each box. Criteria for choosing the sample must be established. Test procedures and acceptance criteria must be developed. The testing will consist of comparing the hard copy documents against the electronic documents stored on the DVD. In comparing the documents checks should be made for quality, readability, and index accuracy. Detailed procedures will be required for conducting the various checks. Specific hardware, software and process details may be required in order to access the documents and index values for testing purposes.